

PATENT APPLICATION
USER INTERFACE FOR DYNAMIC COMPUTING ENVIRONMENT
USING ALLOCATEABLE RESOURCES

Inventor:

Sekaran Nanja
5824 Chambertin Drive
San Jose, CA 95118
Citizen: USA

Assignee:

Design2Deploy, Inc.
2685 Marine Way
Suite 1408
Mountain View, CA 94043

Entity: Small

USER INTERFACE FOR DYNAMIC COMPUTING ENVIRONMENT USING ALLOCATEABLE RESOURCES

CROSS-REFERENCES TO RELATED APPLICATIONS

Sub A15
This application claims priority from U.S. Provisional Application, entitled "Dynamic Computing Environment Using Allocateable Resources" Serial No. 60/228105 and filed on August 24, 2000. This application is also related patent application Serial No. _____ (TBD), entitled System for Configuration of Dynamic Computing Environment Using a Visual Interface, filed September 15, 2000. Both applications and their disclosures are
10 incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

096333-094500
The present invention relates in general to information processing, and more specifically to a system for allowing allocation of resources such as digital
15 processors, networks and software to create a software configured, networked arrangement of processors and processes.

Today, computers are increasingly being used in almost every area of commerce, education, entertainment and productivity. With the growing popularity of the Internet, corporate and campus intranets, home networking and other networks, the trend
20 is to use multiple computers, or processing platforms, to perform tasks and provide services. The networked processors approach is in contrast to traditional approaches of running a single application as a stand-alone application on a single computer.

One approach to networked processing is a "client-server" model whereby a server computer on the Internet is used to transfer information to a client computer.
25 Typically, the client computer is located at an end user's location, such as a personal computer in a user's home. This allows large amounts of information to be stored in, and accessed from, the server computer by many client computers. The client computers can access the server computer simultaneously. Another approach allows a user to obtain portions of executable programs from the server to operate an application program in
30 functional "pieces" or components, on the client computer. For example, a user can run a word-processing program in a client-server mode where the server provides only those portions of the word-processing software to the user's computer on an as-needed basis.

Yet another model is the application service provider (ASP) model. This removes the application still further from the end-user and can involve more processors,

such as servers. The ASP model allows a primary server to host a client-server application, or to host any type of data-processing resource such as a database, user interface, program component, data object, etc. The application can appear to the client as being hosted by the primary server when it is actually being provided by one or more
5 other servers. The other servers can provide the application, or components, by having the client directly access the other server, or having the client access the other server through the primary server.

Still other models are possible. Any arbitrary arrangement of processors can be used to provide a service or function to an end user, or to another device such as
10 another processor, consumer electronic device, etc. Examples of such systems are groups of computers used to perform a large mathematical task, such as rendering frames of a video sequence or performing a simulation. Systems that use many processors to exchange information, such as email systems, multicasting programs, voice-over-IP communications, etc., can potentially require hundreds, or even thousands or more,
15 networked processors.

Software is necessary to operate each of these models. Each of the systems described above will likely use multiple software processes on each processing platform. Most likely, the software processes are designed by different software manufacturers. Although there are standard specifications to design compatible software,
20 such software is still designed largely independently of other manufacturers' software. For example, a user may be running an operating system and web browser designed by different companies. Additional processes may be concurrently executing on the user's computer.

Thus, many independently-designed hardware devices and software
25 processes must be coordinated in order for the overall networked system to operate correctly and efficiently. This complexity is multiplied by the number of users that the system is designed to support. For example, where the system is an Internet system it is common for the user base to be in the hundreds of thousands, or millions. There may be hundreds or thousands (or more) simultaneous users operating client processing systems
30 that require tens, or even hundreds, of servers for adequate bandwidth. Different servers in the networked system may perform different functions such as web page serving, database maintenance, caching, etc.

This complexity further manifests itself in another aspect. In the past, software applications used to be monolithic. That is, the application was self-contained

005760 " 2529950

within a box with no interaction with other applications. However, this paradigm is no longer true. With Internet and e-commerce applications, for example, a client application is configured to talk to a remotely located server application such as eBay.com®, or Amazon.com®, for example. Typically, e-commerce applications define a three-tier architecture which includes: (1) client; (2) server; and (3) database. Within this architecture, the client may be running an Intel® box with a Microsoft Windows 86XX® environment, the server may be running a Sun Microsystems® box running a an iPlanet™ application, the database may be Sun Microsystems® box running Oracle®, for example. Therefore, this three-tiered architecture increases the complexity and cost of many software applications. In one aspect, the complexity of today's computing environments increases computing costs, and extends the time for a new product to reach the market.

A major problem arises where companies need to design and test software and hardware for use in such large and complex systems. Because today's applications are multi-tiered, it is no longer sufficient to test an application within a box. Not only must compatible hardware be selected, obtained, connected and tested; but each piece of software must be selected in accordance with hardware requirements, obtained (usually by complex multi-licensing arrangements), installed, configured and tested. Finally, in order to test the system at anywhere near operating capacity, many human testers may need to be hired to analyze and exercise the system. Often, the task of adequately testing such systems is the single largest expense in designing and developing software. On the other hand, such testing is vitally important to ensure the reliability and success of new software and hardware.

Thus, it is desirable to provide a system and an associated user interface for overcoming the aforementioned problems and which provides advantages over the prior art.

SUMMARY OF THE INVENTION

The present invention discloses user interface for creating a dynamic computing environment using allocateable resources. The interface enables the fast, efficient selection and configuration of processing resources for the computing environment. The resources are fully selectable and allocable by a system architect. In a first embodiment, a primary company, Design2Deploy, Inc.® provides the ability for a customer or system architect to design a system by allocating resources and specifying

how the resources are to be used. The system architect may create a computing environment from a remotely-accessible user interface such as a web page on the Internet. Thus, the system architect can create, modify and operate the environment from anywhere in the world.

5 One embodiment of the present invention provides a method for allocating processing resources, the method using a processor coupled to a display device and to a user input device. The method includes, displaying a list of processing resources on the display device; accepting signals from the user input device to indicate the configuration of at least a portion of the processing resources; and configuring the selected processing
10 resource.

 An alternate embodiment of the present invention discloses a system for providing configurable resources to achieve a processing environment. The system includes a configurable communication link; a plurality of processing devices coupled to the communication link; and a plurality of software programs coupled to the processing
15 devices. A further embodiment of the present invention discloses a method for creating a computing environment by using a computer user interface. The computer user interface coupled to a display screen and to an input device for generating signals in response to interactions of a user. The method includes, accepting a first signal from the input device which enables the user to specify a type of operating system for use in the computing
20 environment; accepting a second signal from the input device which enables the user to specify a type of processor for use within the computing environment; activating an operating system of the specified type to run in the computing environment; and activating a processor of the specified type to run in the computing environment.

 Advantageously, the present invention facilitates the creation of a
25 computing environment on-demand by the customer, and since the computing environment, in turn, can be coupled over networks, including the Internet, this approach eliminates the cost of shipping hardware and software.

BRIEF DESCRIPTION OF THE DRAWINGS

30 Fig. 1 is a block diagram of a system for allocating processing resources according to the present invention.

Fig. 2 is a block diagram of system having a tier architecture for allocating processing resources according to the present invention.

Fig. 3 is an exemplary block diagram of a system showing exemplary service layers in accordance with the present invention.

5 Fig. 4 is a schematic block diagram of a system for providing various services to client network in a secure and distributed environment.

Fig. 5 is a flow chart for describing the basic steps performed by a server system to allocate system resources in response to user interaction according to the present invention.

10

DETAILED DESCRIPTION OF THE DRAWINGS

Detailed descriptions of the embodiments are provided herein. It is to be understood, however, that the present invention may be embodied in various forms. Therefore, specific details disclosed herein are not to be interpreted as limiting, but rather
15 as a basis for the claims and as a representative basis for teaching one skilled in the art to employ the present invention in virtually any appropriately detailed system, structure or manner.

The present invention allows fast, efficient selection and configuration of processing networks. The processing network is referred to as a system including
20 "resources." A system resource is any hardware, software or communication components in the system. For example, discrete hardware devices include processing platforms such as computers or processors, mobile/laptop computers, embedded computing devices, hand-held computers, personal digital assistants, point-of-sale terminals, smart-card devices, databases, storage devices, data transmission and routing
25 hardware etc., without limitation. Additionally, computer peripherals such as monitors, input/output devices, disk drives, manufacturing devices, or any device capable of responding to, handling, transferring or interacting with digital data are also potential resources. Software, or any other form of instruction, is executed by processors in the system and is a type of resource. Finally, communication resources are also part of the
30 system such as a digital network's hardware, the network configuration and topology, and network control as provided by software or hardware.

The resources are fully selectable and allocable by a system architect. In a preferred embodiment, a primary company, Design2Deploy, Inc.® provides the ability for a system architect to design a system by allocating resources and specifying how the

resources are to be used. The system architect can be an individual, corporate entity, etc. The system is referred to as an "environment" and the primary company is referred to as an Environment Service Provider (ESP), while the system architect is referred to as the "customer." The primary company obtains revenue for providing the resources and the tools to easily select, allocate, configure and run the environment.

Note that the preferred embodiment allows such fast allocation and configuration of resources that different environments can be created from the same resources within minutes, or even seconds. This allows "time sharing" of overall resources so that a first environment can be "alive" for a daily two-hour slot, followed by second, third and fourth environments being instantly created for the next four hours for three different customers, and so on.

A preferred embodiment allows customers to create a computing environment from a remotely-accessible user interface such as a web page on the Internet. Thus, the customer can create, modify and operate the environment from anywhere in the world. Since the resources, in turn, can be coupled over networks, including the Internet, this approach eliminates the cost of shipping hardware and software. Designers, programmers, testers or other personnel using an environment can, similarly, be located anywhere in the world so that labor costs can be optimized.

The configuration of environments is automatic. For example, a customer can request a web-site simulator using 12 web-page servers on a Microsoft® NT platform, 2 disk arrays at a specific bandwidth and storage capacity, 2 caching servers and 200 clients running Netscape Navigator™ under Microsoft Windows® 2000 using Pentium III™ processors at under 800 MHz. Such an environment is created automatically. The hardware is actually, or virtually, coupled, the software is automatically installed, and the system can be made available to 200 human testers to operate the browsers. Alternatively, testing software can be used to simulate keystrokes and mouse clicks to perform the testing role. A further understanding of embodiments of the present invention will be gained with reference to the diagrams and descriptions which follow.

System Overview

A product named D2D™ is offered by a primary company, Design2Deploy, Inc.® of Mountain View, California.

Fig. 1 is a block diagram of D2D system 100 for allocating processing resources according to the present invention. In Fig. 1, among other components, system 100 comprises client 142, a computer network such as the Internet 102 and server system 144 which includes various components such as Secure Socket Layer (SSL) 104 and 110, for example. In a first embodiment, server system 144 is a VCE (virtual computing environment), wherein client 142 uses the Internet 102 to interact with server system 144 for the purpose of allocating resources, described below.

Server system 144 may communicate with client 142 using RFB (remote frame buffer) protocol. Other devices and software for facilitating network communications may be employed. In a preferred embodiment, client 142 interacts via SSL 104 and 110, which provide access control so that a desired level of security can be maintained. System 100 further includes a router firewall 106 and Virtual Network Computing (VNC) viewer 112 also used to monitor and control data and to provide security.

Other components of system 100 include an Ethernet switch 128 which allows different hardware components to be flexibly interconnected. For example, Linux boxes 136, Windows Boxes 138, Host Central Processing Units (CPUs) 122, Configuration Host CPUs 130, Domain Name Server (DNS) farm 140 and Redundant Array of Inexpensive Disks (RAID) arrays 124, 126, 132 and 134, can be selected and interconnected. Linux Boxes 136, Windows Boxes 138 processing units which are selectable by users of system 100 and are collectively referred to as "user machines".

Although not shown, one of ordinary skill in the art will realize that the "user machines" may include various machine and device types. For example, such devices may include mobile/laptop computers, embedded computing devices, hand-held computers, personal digital assistants, point-of-sale terminals and smart-card devices without limitation. CPU 122 runs the storage daemon, stores user data and clean configurations. Although not shown, CPU 122 may be connected behind the firewall 120. System 100 further contains other resources such as Application server farm 108, Extended Java Bean (EJB) server 114 and web server farm 118 which are interconnectable via Ethernet switch 128 through Ethernet firewall 120.

Customers' testers or other personnel can interact with the environment through SSL 110 and Web switch 116 to, for example, run client software such as a web browser, application, etc. Note that the selection and interconnection of these devices, along with other degrees of control of these devices as described in the attached

documents, allows a customer to set up many sizes and types of processing architectures and environments.

It is possible to use widely-varying components, or resources, than those shown in Fig. 1 to adequately implement the invention. The specific hardware and interconnections of Fig. 1 are but one example of an architecture for providing allocable resources. Although not shown, system 100 need not include internet 102. In an exemplary embodiment, system 100 comprises a collection of one or wired or wireless networks which are not connected to the Internet. In a first embodiment, the present invention permits "virtual" hosting of computing environments. As used herein, the term "virtual" specifies that neither the requisite devices nor the network is physically accessible to users. Further, in accordance with this embodiment, the hosting process may be initiated or terminated by users at will.

User Interface

In operation, client 142 wishing to allocate resources or create a computing environment connects to a web site served up by server system 144 and provided by Design2Deploy, Inc. System 144 serves up a number of user interfaces to configure the desired resources, as illustrated with reference to the following tables. In a first embodiment, the user interfaces include forms, or other data-entry fields which prompts client 142 to fill in the forms to obtain the desired information. [Add GUI interfaces] Table I below illustrates the contents of a main user interface which is served up upon a request by client 142 for access to system 144.

Main page	
1.	Create a new account
2.	Log into an account

Table I

As shown, option 2 permits client 142 to log into an account that was previously created. System 144 performs username and password authentication to enable client 142 to view the account. It should be noted that system 144 allows each

client to access only that clients' account information. Clients are not free to navigate into another clients web environment to view or allocate resources for another account. When selected, option 1 enables a user interface for creating a new account as showing in Table II, below. The client who created the account is identified as the "owner" of the account. Preferably, each account is associated with only one owner only whom is authorized to change the licensing and billing information for the account.

Other criteria for creating a new account includes the specification of a "licensing plan" by client 142. A "licensing plan" as used herein describes the type and scope of services offered by system 144. For example, a licensing plan "A" may include a selection of five Pentium III processors running on a Microsoft Windows NT platform, 50 client-server licenses and 4 disk arrays at a specific bandwidth and capacity.

Client 142 must also specify billing information when creating a new account. Such information can be an account number or a credit card number, for example. Table II illustrates the content of the user interface for creating a new account.

Create a new account	
Enter new account name:	<accountName>
Enter login name:	<loginName>
Enter password:	<password>
Retype password:	<password2>
Enter your name:	<name>
Enter email address:	<email>
Enter licensing plan:	<licensePlan>
Enter billing info:	<billingInfo>

Table II

Following the creation of a new account, a "laboratory" configuration user interface is served up as illustrated in Table III, below.

Lab page	
1. <u>Edit licensing info</u>	(owner only)
2. <u>Edit billing info</u>	(owner only)

3. <u>Show billing info</u>	
4. <u>Add a user</u>	<i>(owner and administrators only)</i>
5. <u>Delete a user</u>	<i>(owner and administrators only)</i>
6. <u>Change my password</u>	
7. <u>Add a machine</u>	
8. <u>Delete a machine</u>	
9. <u>Add a subnet</u>	
10. <u>Delete a subnet</u>	
11. <u>Open a window on a machine</u>	
12. <u>Shutdown</u>	

Table III

It should be observed that options 1 and 2 are reserved for owners only while options 4 and 5 are accessible to both owners and administrators. Option 1 in table III, allows editing and modification of licensing information as proves necessary. In one embodiment, an error message is generated when client 142 attempts to modify the licensing information in a manner that is inconsistent with the terms of the current laboratory.

Option 3 in Table III, enables client 142 to view information related to a current or past bill. This information may include the number of CPU hours of active machines, the limits to the number of machines allowed and the total amount, for example.

Option 4, Table III allows client 142 to add a new user. As shown below, Table IV illustrates the content of a user interface for adding new users. System 144 restricts access to this portion of the system to only owners and administrators. An administrator as used herein, has all user privileges and can modify/delete both user and administrator accounts.

Add a new user	
Enter login name:	<LoginName>
Enter password:	<password>
Retype password:	<password2>

Enter user's name:	<name>
Enter user's email address:	<email>
Should user be an administrator? [y/n]:	<isAdmin>

Table IV

Option 5 in Table III enables users to be deleted. As in option 4 of table III, access is limited to only owners and administrators. Option 6 enables password modification by client 142.

5 When option 7 is selected, a daughter user interface for adding a machine is presented to client 142, as shown in table V below. Herein lies a first advantage of the present invention. Option 7 allows client 142 to instantaneously create a configurable, secure, and distributed computer environment in accordance with the present invention. Client 142 may create or configure any type of computer environment which includes a
10 desired selection of hardware, software and applications. Advantageously, where client 142 desires to learn about an operating system for example, client 142 may configure a Sun Microsystems® box with the desired operating system platform, such as a Microsoft Windows® environment (80, 95, 98, 2000).

 In a similar fashion, for many users, and developers in particular, this is an
15 invaluable tool. A developer may wish to ascertain the functionality of software in multiple computer environments. The computer environments used for testing must be clean, in the sense that the environment has only original software which are not modified, e.g., by later installed applications, for example. With the present invention, operating systems such as windows 95®, 2000®, Linux 6.2® and solaris® for example,
20 are installable to create a clean computer environment for testing purposes.

 In this manner, the developer may test code using the present embodiment without the disadvantages associated with using a tainted computer environment. Many other benefits of the present invention will become apparent to one of ordinary skill in the art with reference to descriptions and drawings which follow. As shown in exemplary
25 table V, a CPU type, OS type, a subnet and applications used are selectable by client 142 to allocate resources as desired.

Add a new machine

Enter CPU type:	<cpuType>
Enter OS type:	<osType>
Enter owning subnet:	<subnet>
Is gateway for subnet [y/n]:	<isGateway>
Enter hostname:	<hostname> (optional)
Enter applications used:	<apps>

Table V

Option 8, in Table III enables a named machine to be deleted. For example, where a CPU type and an operating system for a particular host name were previously configured, option 8 allows client 142 to delete the particular machine. Some safeguards are built into the system to prevent deletion of a host machine if other client's have open windows on the machine. This rule, however, is inapplicable to owners and system administrators who can shut down such active machines.

Add a new subnet	
Enter owning subnet:	<parentSubnet>
Enter subnet name:	<subnetName> (optional)

Table VI

Option 9 in Table III is for the purpose of adding a new subnet. Table VI illustrates the content of the usual interface for adding a new subnet. A "subnet" as used herein is defined as all of the machines at one geographic location, in one building, or on the same local area network (LAN). Client 142 must specify which subnet the selected subnet is nested within. The computing environment itself may be specified for this field, and it may be indicated that the subnet is in the root subnet and is not connected any other subnets. As a default, server system 144 generates a name for the subnet if client 142 fails to specify a name.

Delete a subnet

Enter subnet name:	<subnetName>
--------------------	--------------

Table VII

Option 10 in Table III, enables the deletion of a previously selected subnet. System 144 prevents the deletion of a subnet that contains active machines. The user interface for this option is illustrated above in Table VII.

5

Open a window on a machine	
Enter machine name:	<hostname>

Table VIII

Selection of option 11 in table III pops up a different window showing the desktop for the requested machine, if a machine has been activated. Otherwise, if a requested machine has not been activated, it is activated and the window is opened. In addition, all gateway machines of enclosing subnets for this machine will also be activated. Table VIII shows the contents of the "open a window on a machine" user interface. When clicked, option 12, table III shuts down one or more previously configured machines. Table IX below illustrates the user interface for shutting down a machine. In one embodiment, the shutdown user interface may provide capabilities to shut down entire subnets or the entire laboratory.

10

15

Shutdown	
Enter machine name:	<hostname>

Table IX

Table X illustrates a login user interface through which client 142, if previously registered, may access system 144. Access is authorized only if client 142 is any one of an owner, administrator, and a user. Administrators have all the privileges that users have and can delete user and administrator accounts. Owners have all of the privileges that administrators do and can change licensing and billing information.

20

Login Page	
Enter account name:	<accountName>
Enter login name:	<LoginName>
Enter password:	<password>

Table X

Table X illustrates a login user interface for logging into system 144 for a previously registered client. Access to system 144 is permitted only if client 142 is any one of an owner, administrator, and a user. Administrators have all of the privileges that users have and can delete user and administrator accounts. Owners have all of the privileges that administrators do. A “gateway machine” is defined as a machine that acts as a network point to a network or subnetwork. In one embodiment, the gateway machine may act as a proxy server or a firewall server. Server system 144 may fail to activate a machine or its gateway in some circumstances. Such circumstances include insufficient hardware resources, insufficient operating systems or not enough application licenses for the requested resources. In this case, system 144 generates and displays the appropriate error.

N-Tier Architecture

Fig. 2 is a block diagram of D2D System 200 having a tier architecture according to the present invention.

In Fig. 2, system 200 comprises a number of tiers, namely switch tier 203, web 205 and usability tier 207. Other tiers include middleware tier 209, application logic tier 211 and data center tier 213. One or more of the tiers are implementable using software (dedicated custom or third-party), or hardware or a combination of both. Switch tier 203 includes a router 217 for routing data packets through the network, an FW () and a balancer 219 for balancing internet protocol loading. Router 217 may be a Cisco 7200 Series™ router available from Cisco, Inc. ® or other comparable type routers. Web tier 205 comprises a plurality of web servers such as a Sun Microsystems® box running an iPlanet™ web server, enterprise edition 4.1 for example, or other comparable type web servers. Usability tier 207 provides various services including load balancing, building, session managing, SSL and fault tolerance.

Middleware tier 209 contains one or more application servers 221, 223 and a module 225 for implementing look up event and services. Application server 221 may be DEA web logic server™ running within a Windows NT® platform, for example. Application logic tier 211 provides a variety of operating systems test services such as Windows 227, Linux 229, Unix 231 test services. User management service 233 is implemented within application logic tier 211. Data center tier 311 includes various operating system platforms and processors, also selectable by the user. Data center tier 311 in one embodiment is removed from the application logic tier 211. Although not shown, one of ordinary skill in the art will realize that one or more of the aforementioned tiers and components therein are implementable using third party providers, dedicated custom modules or software and hardware or a combination thereof.

Fig. 3 is a block diagram of D2D System 300 showing exemplary service layers in accordance with the present invention. As shown, system 300 comprises client 301, communicatively coupled through the Internet 307, to server system 343.

In Fig. 3, firewalls 303 and 305 maintain security within a client network and the server system 343, respectively. In a simple configuration shown in Fig. 3, server system 343 comprises web server 309, application server 311 and a plurality of databases 345. Web server 309 includes JSPs (Java server pages) 349 which coordinate with servlets 347 to generate the user interfaces seen by the user. Although the user interfaces receive information (e.g. commands to open windows or to shut down machines), they may display information such as account and user information, machine configuration information, acknowledgments, error and state information, for example. These functionalities are accomplished, at least in part, through communication with a session manager 331. In one embodiment, the communication (with session manager 331) uses XML (extensible markup language) or other comparable low-level code.

As shown in Fig. 3, application server 343 functions to run a plurality of services namely, a user manager 329, a session manager 331, a test service manager 333 and a storage service manager 335. Others include a CPU manager 337, DHCP manager 339 and license pool manager 341. Although not shown, each of the aforementioned services are operable in a distributed environment having one or more servers and computing systems.

User manager 329 is responsible for holding all user account information specific to single user. Such information includes the login name, password, user name and email address for the user. It will be apparent to one of ordinary skill in the art that

additional information fields may be added to a user account as proves necessary. Such fields may be the machine login name and the user identification, for example. User beans are uniquely identified using the tuples: (account name, login name).

Session manager 331 is the intermediary between client 142 (Fig. 1) and server system

5 343. All requests from and responses to client 142 involve session manager 331. In one embodiment, session manager 331 solely functions to route these messages and replies between the appropriate managers and client 142. Therefore, session manager 331 performs little or no computations and has no state. Implementation of session manager 331 may be achieved using a Java session bean, in one embodiment. A "bean" as herein
10 used refers to a component which is a reusable program building block that can be combined with other components in the same or other computers in a distributed network to form an application.

Test services manager 333 functions to map physical machines to a laboratory configuration that indicates a user's virtual selection of allocateable resources.

15 Test services manager 333 acts as an intermediary between various sub-managers responsible for setting up, shutting down and accessing physical machines and a configuration manager (not shown). The configuration manager functions to **need functionality**. Like the user manager 329, test services manager 333 has no state, and its primary purpose is to coordinate operations between other managers. Additionally, test
20 services manager 333 may roll back one or more previously completed services if an invoked service fails. It is also responsible for updating the state of all machines and subnets. However, it should be noted that a machine's status will only be updated to up or down when test services manager 333 is fully completed.

In an exemplary embodiment, test services manager 333 delegates certain
25 operations initiated by session manager 331 to other sub-managers. The operations are (1) startup of a machine; (2) shutdown of a machine; (3) shutdown of the entire lab; and (4) opening a machine window.

The sub managers to which these operations are delegated include a startup service sub manager (not shown) and a storage service sub manager (not shown).

30 As implied by its name, the startup service sub manager functions to startup and shut down machines. Upon identifying the appropriate machines for starting up or shutting down, it directs CPU manager 337 to execute the task. After a set of machines are started up, for example, the sub manager will return those machines in a vector so that those machines can be shut down if a rollback is needed. The startup service sub manager is

also responsible for keeping track of the number of operating system licenses available in the license manager (described below) and updating the number of running machines within system 343. An addition function of the startup service sub manager is to communicate with DHCP manager 339 to both allocate and free IP addresses for the machines. The storage service sub manager is responsible for mounting and unmounting the shared persistent storage on the machines. This functionality is achieved by communicating and coordinating with storage service manager 335.

Other sub managers to which tasks are delegated by the test service manager are a login service sub manager (not show) and an application service sub manager (not shown) for installing and uninstalling applications on the user selected machines. The applications sub manager keeps in contact with license pool manger 341 to update the number of licenses. The application sub manager will invoke the CPU service manager 337 or an application service manager (depending on the implemented embodiment) to perform the installs and uninstalls. Alternately, neither the application service sub manager or manager is implemented so that all installs/uninstall are performed by the startup service sub manager.

With regard to the login service sub manager, it functions to login a user into a machine and pops a VNC window of the machine's desktop for the user by coordinating with a login manager (not shown). It should be observed that one or more of the aforementioned sub managers are implemented as stateless session beans so that some variables are cached. Additionally, although not shown one of ordinary skill in the art will realize that one or more of the sub managers and associated functionalities may be combined, or implemented as other service layers consistent with the spirit and scope of the present invention.

In Fig. 3, storage service manager 335 functions to keep track of all of the shared storage in a session. Storage service manager 335 allocates storage for a particular laboratory, and mounts/unmounts this storage on all of the machines in the laboratory. Storage manager 335 only interacts directly with the storage service sub manager (of test service manager 333). Storage service manager 335 typically has a daemon running on each machine within system 343 and it will communicate these daemons through a look up service. In one embodiment, the interface between storage service manager 335 and the storage service sub manager is implemented as a session bean.

CPU manager 337 functions to track the machines that are within system 343 as well as to assign one or more of these machines as virtual machines. When the

start up service sub manager requests a machine with a particular CPU and operating system, CPU manager 337 locates such a machine and, if need be, installs the proper operating system on the machine. CPU manager 337 further functions to return a machine into a free pool when released by start-up service. In addition, if an actively
5 used machine crashes, CPU manager 337 informs the start up service manager that such a crash occurred. Each machine in System 343 runs a CPU-based daemon to keep track of the active and running machines. A look up service will be utilized to perform this functionality. It should be noted that the interface between the CPU manager 337 and the rest of D2D system 300 is implemented as a session bean.

10 DHCP manager 339 functions to retrieve and free IP addresses. System 343 allocates an IP address for a machine just before it is started up and frees the address after the machine shuts down.

License pool manager 341 keeps track of the number of operating systems and application licenses utilized by system 343. Each operating system and application
15 license has a maximum number of licenses. License pool manager 341 prevents system 341 from exceeding the number of allocated licenses. Each license in license manager 341 is implemented as an entity bean. As shown in Fig. 3, the aforementioned services utilize one or more databases for storing their associated information. In a first embodiment, these databases utilize a database server running SQL 7.0 database server
20 software by Microsoft and hosts additional databases each serving specific functions. Further, these database architecture contains an underlying ODBC-compliant Microsoft SQL Server database allowing for platform-independent data transfer but may operate with any ODBC compliant database.

In Fig. 3, information database 313 stores user information for user
25 manager service 329, a database 315 stores information relating to sessions 91 for session manager 331, a test configuration database 317 stores associated information relating to test configuration data and is to storage service manager 335. Databases 321, 323 and 325 store applications, user data and CPU and IP address information, respectively, for storage service manager 335. Database 325 is coupled to DHCP manager 339 as well.
30 A database 327 for storing licensing information is communicatively coupled to license pool manger 341.

The primary company Design2Deploy, Inc., of Sunnyvale, Ca. offers various products/services namely zap!Safe™ zap!Test™ service and zap!Deploy™ which will be described with reference to Fig. 4, below.

Fig. 4 is a schematic block diagram of D2D system 400 for providing various services to client network 401 in a secure and distributed environment. A VPN (virtual private network) client 403 within client network 401 may request zap!Safe™ storage 407 from a VPN server 405 (of system 400). Storage 407 offers persistent storage capabilities within a VPN 409 that are accessible from all testing and deployment configurations. In one embodiment, storage 407 offers a single site to store all mission critical data and applications for an extended period of time, and features full back-up, disaster recovery and encryption facilities.

A number of other services are offered by system 400 namely zap!Test service 411 and service 413. Corporate internet 415 may request service 411 which offers click and go testing configurations for various hardware, software and operating systems, and zap!Deploy™ 413 services for enabling on-the-fly (beta site) deployment and staging capabilities for various operating system platforms.

Fig. 5 is a flow chart for describing the basic steps performed by server system 144 to allocate system resources in response to user interaction.

At step 502, server system 144 awaits user input. In one embodiment, user input is via a user input device coupled to a processor and display device of a computer. The user operates the computer to access server system 144 over the Internet. The computer can accept signals from the user input device to indicate the preferences of the user.

Server system 144 receives the signals via web page information that includes forms, or other data-entry fields, and prompts the user to fill the form to obtain the desired information. Although not shown, the web interface may be solely graphical, textual or a combination of both. For example [insert new config?] In Fig. 5, the flowchart is entered at step 502, when the user desires to allocate system resources.

At step 504, the user selects a "login option". It is assumed that a valid account has previously been opened for the USER1, under an account named ACCOUNT1, for example. Next, the user provides the user's information, namely, USER1, ACCOUNT1, and a password. Session manager 331, system server 144 creates an instance of session manager 331 for USER1. Session manager 331 locates the user bean named USER1 under ACCOUNT1 and authenticates the user identification, account and password.

Next, if USER1 has previously configured a laboratory e.g., LAB1 for ACCOUNT, it is retrieved. Otherwise, USER1 is presented with a "laboratory

configuration” page. Both text and visual based examples of the “laboratory configuration” page are shown in Fig. _.

At step 506, USER1 creates a subnet named SUBNET1 with no parent subnet. To implement this configuration, session manager 331 directs the configuration
5 manager to create SUBNET1 under ACCOUNT1 and LAB1, which is assigned an identification number of 1. At this point, the status of SUBNET1 is down.

At step 508, the processing resources are configured. On clicking “add a machine” on the menu option, USER1 creates a processing resource such as a x86 Linux machine running a “solitaire 2.0 application, named CLIENT1, for example, and
10 configured under SUBNET1. Session manager 331 directs the configuration manager to create a new machine bean named CLIENT1 under ACCOUNT1 and LAB1. The machine is assigned an x86 CPU, a Linux operating system, an identification number of 2, a down status and its subnet assigned 1.

At decision box 510, USER1 may configure or add as many machines, and
15 operating systems and software as desired. If for example, USER1 wishes to configure a server for the computing environment, USER1 clicks on “add a machine” again. USER creates a solaris machine named SERVER1, under SUBNET1 with no applications. To implement this configuration, session manager 331 directs the configuration manager to create a new bean for SERVER1 under account ACCOUNT1 and LAB1. A SPARC
20 CPU, a solaris operating system, machine identification number 3, down status and machine subnet 1 are assigned.

At step 512, USER1 may open a window on the machines (e.g. CLIENT1) which were previously configured by selecting “open a machine window”. In response, session manager 331 creates a new instance of test service manager 333 (Fig. 3) and
25 invokes the startup service sub manager (not shown) within test service manager 333. Test service manager 333 obtains the status of CLIENT1, thereafter invokes the startup sub manager to bring up CLIENT1. The startup sub manager creates an instance of the startup service and invokes CLIENT1.

In addition, the startup service performs the following operations; (1)
30 changes the status of the machine to “starting up”; (2) requests a Linux license from license pool manager 321; (3) requests an IP address from DHCP manager 339; (4) directs CPU manager 337 to obtain an x86 machine running Linux and to assign the IP address, and host name CLIENT1. In response, CPU manager 337 creates the machine, and returns a CPU key identifying the machine for storage in the machine’s configuration

bean. Thereafter, startup service returns with a vector of all machines that were started up.

5 Additionally, in order to open the window for CLIENT1, the startup service sub manager creates an instance of the storage service and invokes it on all of the running machines (currently CLIENT1). The storage service checks whether the shared storage for LAB1 has been allocated. If not, storage service invokes the storage service manager 335 to create a persistent shared storage pool for this LAB1. Storage service manager 335 creates this storage pool and returns a key to the storage pool, which is stored within LAB1. Thereafter, the storage service invokes the storage service manager 10 335 to mount the storage pool with the given key on CLIENT1. Storage service manager 335 responds by mounting the storage pool.

15 A further step which must be taken to open the window for CLIENT1 includes, creating an instance of application service and invoking it on all started up machines (i.e. CLIENT1). Next, the application service sub manager requests an application license for "solitaire 2.0" and installs the application after the license is granted. After "Solitaire 2.0" is installed, the status of CLIENT1 is switched to "up." Because the machines (CLIENT1) are now up and running, the startup service sub manager opens a window CLIENT1 by creating an instance of the login service sub manager.

20 The login service sub manager is primarily responsible for performing the following functions: (1) checking whether ACCOUNT1 contains a valid Unix login and user ID; (2) checks whether USER1 is a valid Unix login name and that it is not being used by any other user within server system 144; (3) if USER1 exists, the login server generates a second valid username USER2; and (4) generates the user ID that is unique 25 within the service system 144. Thereafter, the login service sub manager determines a GID (_ identification) from the storage pool key, and generates a UID (user identification) that is unique within the server system 144.

30 The login service sub manager determines the GID from the storage pool key, and directs the login manager to log in USER2 with the given UID and GID into the machine CLIENT1. The login manager checks whether user USER2 exists as a user on CLIENT1. If USER2 does not exist, USER2 is created with the given UID, GID and with no password. Next, The login manager checks whether a home directory for USER2 was allocated in the shared pool.

If unallocated, a home directory and the shared storage pool are created for USER2. The password file for USER2 is updated to use this home directory as USER2's home directory. Next, the login manager _____ to USER2, starts the VNC server daemon and returns the URL (uniform resource locator) to access the VNC server. In turn, test service manager 333 returns the URL to access the VNC server.

In this manner, the present invention permits JSP page 349 or servlet 347 (Fig. 3) to pop up a daughter window in the browser with the URL address. In the daughter window, the desktop for CLIENT1 is seen by USER1 such that USER1 may interact with CLIENT1 to carry out one or more tasks such as code testing (for example), in accordance with the present invention.

Although the present invention has been discussed with respect to specific embodiments, one of ordinary skill in the art will realize that these embodiments are merely illustrative, and not restrictive, of the invention. The scope of the invention is to be determined solely by the appended claims.